DOCUMENT RESUME

ED 420 679 TM 028 354

AUTHOR Timm, Neil H.
TITLE Analyzing Interactions in Unbalanced Two-way Designs.
PUB DATE 1998-00-00
NOTE 18p.
PUB TYPE Reports - Evaluative (142)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Interaction; Research Methodology; Researchers
IDENTIFIERS *Factorial Designs; *Unbalanced Designs; Weighting (Statistical)

ABSTRACT

This paper considered several issues with the analysis and interpretation of interactions in unbalanced factorial designs. The effect of design weights on the interaction parameters of factorial designs and an approach for the analysis of interactions using finite intersection tests is discussed. The example of a researcher designing a study to analyze the interaction of two treatment combinations is used. In implementing the study, an unbalanced nonorthogonal design resulted. To analyze this data, one researcher in the example used equal cell weights and another used proportional cell or balanced weights. It is shown that the interaction parameters and their corresponding estimates depend on the weighting scheme selected. Implications of these findings are discussed. (Contains 38 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 420 679

Analyzing Interactions in Unbalanced Two-way Designs

Neil H. Timm

University of Pittsburgh

Key Words: Factorial design, design weights, simultaneous inference, finite intersection tests, nonorthogonal

Abstract

This paper considers several issues with the analysis and interpretation of interactions in unbalanced factorial designs. The effect of design weights on the interaction parameters in factorial designs and an approach for the analysis of interactions using finite intersection tests is discussed.

To motivate and illustrate the issues discussed in this article, we begin with a simple example. Consider the problem of a researcher designing a study to analyze the interaction of two treatment combinations (factors A and B) where factor A has three levels and factor B has four levels. Following Kirk (1995, Chapter 9), a completely randomized factorial design, represented as CRF-ab where a = 3 and b = 4. Kirk (1995, p. 422) and most other authors of texts in classical experimental design in the social and behavioral sciences, biostatistics and statistics recommend that the researcher (1) assume a balanced design with an equal number of observations per cell, (2) establish the magnitude of the interactions (γ_{ij}) one wishes to detect with the desired power for the overall test of the interaction, with an estimate of common within subject variance, and (3) determine the sampling plan for the study. Alternatively, following Cohen (1988), an estimate of the effect size for the test of interaction is specified to again determine the equal number of observations per cell.

Following the procedure discussed by Kirk (1995, p. 401), the power analysis of a researcher showed that exactly 2.75 subjects per cell or $N = 33$ subjects were required for a study to maintain a 0.80 level of power to reject the overall test of interaction (AB). While three observations per cell yielded a slightly higher power, the researcher knew that a design with an equal number of observations per cell was "easier" to analyze; hence, the researcher decided on a sample size of $N = 36$ subjects.

A second researcher knew that factors A and B contained mutually exclusive and exhaustive treatment levels associated with the two factors. Furthermore, this researcher knew that the number of observations associated with the levels of factor A in the population were equal and that the number associated with a levels of factor B were proportional in the population with fixed ratios 2:3:2:4. Using this information, the researcher decided on the sampling plan shown in Table 1, an orthogonal design with proportional cell frequencies.

TM028354

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Neil Timm

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Table 1 Design Weights W_{ij}

		Factor B				
		B_1	B_2	B_3	B_4	
Factor A	A_1	2	3	2	4	11
	A_2	2	3	2	4	11
	A_3	2	3	2	4	11
		6	9	6	12	33

Thus, we have two design strategies: an "equal" weight design and a "proportional weight" or "balanced" design.

Collecting the data for the study and performing the experiment with $N = 36$ subjects, several data values were lost and the data realized for the experiment were those shown in Table 2, from Overall and Spiegel (1969); an unbalanced or nonorthogonal design. The second researcher began his study with $N = 33$ subjects according to the sampling plan given in Table 1; however, one subject left the experiment and some of the subjects were incorrectly assigned to the treatment combinations. Again, the data shown in Table 2 were realized. Finding yourself in this situation, how would you analyze the data in Table 2? Should you employ equal weights, balanced weights, unequal weights or the cell frequencies as weights - "sample" weights? Does the selection of weights make a difference for testing the interaction hypothesis or estimating the interaction effects?

Table 2 Overall and Spiegel Data

		Factor B			
		B_1	B_2	B_3	B_4
Factor A	A_1	61 73 52	79 65 81	43 35	56 25 19 35
	A_2	42 53	37 32 50	87 81 65	72 84
	A_3	96 81 92	45 37	75 59	98 77 91

Before answering this question, let us see what others say about the analysis of a two-way CRF-34 analysis of variance (ANOVA) with fixed effects and unbalanced data as shown in Table 2. While there has

been considerable discussion of this situation in the social and behavioral sciences by many authors, for example, Appelbaum and Cramer (1974), Carlson and Timm (1974), Cramer and Appelbaum (1980), Keren and Lewis (1977), Lewis and Keren (1977), O'Brien (1976), Overall and Spiegel (1969, 1973), Overall, Spiegel, and Cohen (1975), Rawlings (1972, 1973), Searle (1995), and Timm and Carlson (1975), among others, there is not total agreement among the authors on the analysis. Most authors agree that selecting sample weights, weighting by cell frequencies, is incorrect unless the unbalanced number of observations observed in all cells would result in proportional patterns over several replications of the experiment. Following Scheffé (1959, p. 93), others might say that the test of interaction does not depend on the system of weights. Only the tests of main effects depend on the weights. This is correct; however, the interaction parameters γ_{ij} and their estimators, $\hat{\gamma}_{ij}$, depend on the weights used in the design. This led Arnold (1981, pp. 92-100), Davidson and Toporek (1991) and Fujikoshi (1993) to assert that the researcher first establish the system of design weights or restrictions on the model. Following the establishment of design weights or model restrictions, one next determines the sampling plan for the design based on power considerations. Unfortunately, there is no "optimal" strategy in the selection of the system of design weights, Fujikoshi (1993).

Returning to our example, we would not weight by the sample cell frequencies or use weighted averages of cell means since the n_{ij} are not random. The first researcher would select an equal cell weight analysis which associates weights $(1/a)$ to each row of factor A and $(1/b)$ to each row of factor B, provided the treatment levels are mutually exclusive and exhaustive for both factors as suggested by Carlson and Timm (1974), among others. For the second researcher, proportional or balanced weights should be used in the analysis since they reflect the proportion of the population in the j^{th} column. Does the analysis make a difference? Yes, before we look at four approaches and the difference in the analysis, we introduce some notation.

The Two-Way ANOVA Model

For our study we have a CRF-ab design; the full rank cell means linear model is:

$$y_{ijk} = \mu_{ij} + e_{ijk}, \qquad i = 1,2,\ldots,a; \quad j = 1,2,\ldots,b; \quad k = 1,2,\ldots,n_{ij}, \tag{1}$$

where the e_{ijk} are independent, normal, random errors with mean zero and common variance, σ^2. Furthermore, we assume that the cell means μ_{ij} have the following form:

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}, \tag{2}$$

where μ is a general constant, α_i is the effect due to the i^{th} level of factor A, β_j is the effect due to the j^{th} level of factor B and γ_{ij} represents the effect of the interaction of row i and column j. To associate meaning to the parameters in (2), to make them uniquely estimable, one imposes a set of restrictions on the parameters:

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0, \text{ or} \tag{3}$$

$$\sum_i U_i \alpha_i = \sum_j V_j \beta_j = \sum_i U_i \gamma_{ij} = \sum_j V_j \gamma_{ij} = 0 \tag{4}$$

where the weights U_i and V_j are nonnegative such that the $\sum_i U > 0$ and $\sum_j V_j > 0$, or

$$\sum_i W_{i+} \alpha_i = \sum_j W_{+j} \beta_j = \sum_i W_{ij} \gamma_{ij} = \sum_j W_{ij} \gamma_{ij} = 0 \tag{5}$$

where the W_{ij} are nonnegative cell weights, $W_{i+} = \sum_j W_{ij}$ and the $W_{+j} = \sum_i W_{ij}$. These systems of weighting schemes or restrictions are called Σ-restrictions, *UV*-restrictions and *W*-restrictions by Fujikoshi (1993).

In (3) we see that selecting $U_i = V_i = 1$ or $W_{ij} = 1$ so that equal weight Σ-restrictions is a special case of the *UV*- or *W*-restrictions, and $W_{ij} = U_i V_j$. If W_{ij} is proportional to the product of U_i and V_j, we say the design is "balanced" with respect to the weighting scheme. In particular, if we let

$$\begin{aligned} &\sum_j W_{ij} = W_{i+} \text{ and } U_i = W_{i+} / W_{++} \\ &\sum_i W_{ij} = W_{+j} \text{ and } V_j = W_{+j} / W_{++} \\ &\sum_{i,j} W_{ij} = W_{++} \text{ and } \sum_i U_i = \sum_j V_j = 1, \end{aligned} \tag{6}$$

then $W_{ij} \propto U_i V_j$ and the scheme is balanced. We may also select $W_{ij} = n_{ij}$ so that $W_{++} = n_{++}$ and other more abstract schemes. In general, *UV*-restrictions are not special cases of *W*-restrictions, unless all U_i and V_j are positive, Fujikoshi (1993). Furthermore, it is not possible to develop simple expressions for all the model parameters μ, α_i, β_j and γ_{ij} for the general *W*-restrictions unless the weights are balanced since solutions for the α_i's and β_j's involve solving the absorbing equations, Searle (1971, p. 297).

For the weighting scheme $W_{ij} = U_i V_j = (1/\mathrm{a})(1/\mathrm{b})$, which ensures positive, balanced U_i and V_j, the estimator of the interaction parameter is

$$\begin{aligned} \hat{\gamma}_{ij} &= y_{ij\cdot} - \sum_j y_{ij\cdot} / \mathrm{a} - \sum_i y_{ij\cdot} / \mathrm{b} + \sum_i \sum_j y_{ij\cdot} / \mathrm{ab} \\ &= y_{ij\cdot} - y_{i\cdot\cdot} - y_{\cdot j\cdot} + y_{\cdot\cdot\cdot} \end{aligned} \tag{7}$$

where $y_{ij\cdot}$ is the cell mean. The "dot" notation is used to represent a simple or unweighted average. For $U_i = n_{i+}$ and $V_j = n_{+j}$ or $W_{ij} = n_{i+} n_{+j}$ so that $W_{ij} > 0$ and balanced, the estimator of the interaction term is

$$\begin{aligned} \hat{\gamma}_{ij} &= y_{ij\cdot} - \sum_j n_{+j} y_{ij\cdot} / n_{++} - \sum_i n_{i+} y_{ij\cdot} / n_{++} - \sum_i \sum_j n_{ij} y_{ij\cdot} / n_{++} \\ &= y_{ij\cdot} - \bar{y}_{i\cdot\cdot} - y_{\cdot j\cdot} + \bar{y}_{\cdot\cdot\cdot} \end{aligned} \tag{8}$$

where the "bar" notation represents a weighted average of cell means. In (7), we have unweighted averages of cell means so that this is often termed the equal weight case. Situation (8) is called the unequal weighted case, since the $\hat{\gamma}_{ij}$ depend on weighted averages of cell means. More generally, suppose we have arbitrary weights where the weighting is specified by (5) with weights $\{W_{ij}\}$ where not all $W_{ij} > 0$ and the W_{ij} are not balanced. Then,

$$\hat{\mu} = \sum_i \sum_j W_{ij} y_{ij} / W_{++}$$
$$\hat{\gamma}_{ij} = y_{ij\cdot} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j \qquad (9)$$

where $\sum_i W_{i+}\alpha_i = 0$, the $\sum_j W_{+j}\beta_j = 0$ and $\hat{\alpha}_i$ and $\hat{\beta}_j$ are solutions to the general absorbing equations, Fujikoshi (1993). If $W_{ij} = n_{ij} = n_{i+}n_{+j} / n_{++}$, we have the case of proportional sampling. This is not the case for our example since $W_{ij} = W_{i+}W_{+j} / W_{++}$ and $W_{ij} \neq n_{ij}$. We call this situation the "balanced" weight case, since $W_{ij} \propto W_{i+}W_{+j}$. For $W_{ij} = n_{ij}$, $\hat{\alpha}_i$ and $\hat{\beta}_j$ do not have simple representations and (8) does not apply.

Using the population cell means μ_{ij}, the corresponding row and column population marginal means are defined:

Row	Column
$\mu_{i\cdot} = \sum_j \mu_{ij} / \mathrm{b}$	$\mu_{\cdot j} = \sum_i \mu_{ij} / \mathrm{a}$
$\bar{\mu}_{i\cdot} = \sum_j n_{ij}\mu_{ij} / n_{i+}$	$\bar{\mu}_{\cdot j} = \sum_i n_{ij}\mu_{ij} / n_{+j}$
$\tilde{\mu}_{i\cdot} = \sum_j W_{ij}\mu_{ij} / W_{i+}$	$\tilde{\mu}_{\cdot j} = \sum_i W_{ij}\mu_{ij} / W_{+j}$.

(10)

For the model given in (1) and all $n_{ij} > 0$, Graybill (1976, p. 560) has shown that the test for no interaction or additivity may be represented in any of the following four equivalent forms:

$$\begin{aligned}
&\text{(a) } H_o\colon \gamma_{ij} = 0 \\
&\text{(b) } H_o\colon \mu_{ij} = \mu + \alpha_i + \beta_j \\
&\text{(c) } H_o\colon \mu_{ij} - \mu_{i'j} - \mu_{ij'} + \mu_{i'j'} = 0 \\
&\text{(d) } H_o\colon \gamma_{ij} - \gamma_{i'j} - \gamma_{ij'} + \gamma_{i'j'} = 0
\end{aligned} \qquad (11)$$

for all subscripts i, i', j and j'.

To test (11), one uses the F-statistic:

$$F = \frac{\sum_{i,j} n_{ij}(y_{ij\cdot} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j)^2 / (\mathrm{a}-1)(\mathrm{b}-1)}{\sum_i \sum_j \sum_{k=1}^{n_{ij}} (y_{ijk} - y_{ij\cdot})^2 / (N - \mathrm{ab})} = \frac{MS_H}{MS_e} . \qquad (12)$$

When H_o as in (11) is true, the F-statistic in (12) has a central F-distribution with $\nu_h = (\mathrm{a}-1)(\mathrm{b}-1)$ and $\nu_e = N - \mathrm{ab}$ degrees of freedom. For a deviation of the statistic in (12), see for example, Scheffé (1959, p. 115), Arnold (1981, p. 96) or Fujikoshi (1993). The test is a uniformly most powerful invariant, unbiased test, Arnold (1981, p. 109).

While the F-test of (11) using (12) is independent of the system of weights (Scheffé, 1959, p. 93; Arnold, 1981, p. 95 and Fujikoshi, 1993, Theorem 3.2, p. 320), if the interaction hypothesis is rejected so that not all the γ_{ij}'s are identically zero, the estimands γ_{ij} depend on the design weights and hence so may the interpretation of their estimates, $\hat{\gamma}_{ij}$.

Analyzing the Two-Way ANOVA

Returning to our example, we analyze the data in Table 2 using the BMDP4V computer package, BMDP(1992). This statistical package allows the researcher to input design weights using the /WEIGHTS paragraph. The parameter EQUAL assigns cell weights $W_{ij} = (1/a)(1/b)$. The option SIZES assigns cell weights $W_{ij} = n_{ij}$, a special case of the *W*-restrictions discussed by Fujikoshi (1993). With the LIST option one may assign weights to each level of the design. Using the weighting $A = 11, 11, 11$ and $B = 6, 9, 6, 12$, we obtain the unequal weighted case. Selecting the weights proportional to cell frequencies, $A = 12, 10, 10$ and $B = 8, 8, 7, 9$, we obtain the "weighted" average case. The unequal weight case yield interactions that have the form given in (8). The parameter matrices for the cell means and interaction parameters are the $a \times b$ matrices:

$$\mathbf{U} = (\mu_{ij}) \text{ and } \Gamma = (\gamma_{ij}) \tag{13}$$

where μ_{ij} is a cell mean and γ_{ij} is an interaction, and $i = 1, 2, 3$ and $j = 1, 2, 3, 4$, for our example.

For the data in Table 2, the estimator of **U** is the matrix of sample cell means $y_{ij\cdot}$; hence,

$$\hat{\mathbf{U}} = (y_{ij\cdot}) = \begin{pmatrix} 62.00 & 75.00 & 39.00 & 33.75 \\ 47.50 & 39.67 & 77.67 & 78.00 \\ 89.67 & 41.00 & 67.00 & 88.67 \end{pmatrix}. \tag{14}$$

From the output of BMDP4V, one may construct the matrix of estimated interactions, $\hat{\Gamma}$. Using estimates of the population marginal means given in (10) and an estimate of the overall mean μ, the interactions $\hat{\gamma}_{ij}$ are obtained using (7) for the "equal" or unweighted weight case. Using (8), we obtain the unequal weight case. BMDP4V solves the general absorbing equations to estimate $\hat{\gamma}_{ij}$ when selecting weights $W_{ij} = n_{ij}$, the "sample" weight case. For the "balance" weight case,

$$\hat{\gamma}_{ij} = y_{ij\cdot} - \hat{\bar{\mu}}_{i\cdot} - \hat{\bar{\mu}}_{\cdot j} + \sum_i \sum_j W_{ij} y_{ij\cdot} / W_{++} \tag{15}$$

where $W_{ij} = W_{i+} W_{+j}$. The twelve interaction estimates $\hat{\gamma}_{ij}$ are the cells in Table 3 for the four weightings: equal, sample, balanced and unequal.

Table 3 3x4 Design, Interactions (γ_{ij}), Marginal Means, Overall Mean

	Equal Weight Case ($W_{ij} = 1$)				$\mu_{i\cdot}$
	4.7500	32.2500	-13.0833	-23.9167	52.4375
	-18.0208	-11.3542	17.3125	12.0625	60.7083
	13.2708	-20.8958	-4.2292	11.8542	71.5833
$\mu_{\cdot j}$	66.3889	51.8889	61.2222	66.8056	61.5763

Sample Weight Case ($W_{ij} = n_{ij}$)

	3.8591	30.6703	-14.3769	-18.7086	
	-19.5090	-13.5312	15.4217	16.6733	
	9.1469	-25.7086	-8.7557	13.8293	
					61.8125
		Balanced Weight Case ($W_{ij} = W_{i+}W_{+j}$)			$\tilde{\mu}_{i.}$
	6.1667	33.6667	-11.6667	-22.5000	51.0909
	-19.1818	-12.5152	16.1516	10.9015	61.9394
	13.0152	-21.1515	-4.4848	11.5985	71.9091
$\tilde{\mu}_{.j}$	66.3889	51.8889	61.2222	66.8056	61.6465
		Unequal Weight Case ($W_{ij} = n_{i+}n_{+j}$)			$\bar{\mu}_{i.}$
	4.7705	30.5518	-11.9482	-22.1045	52.2734
	-18.1748	-13.2269	18.2731	13.7002	60.7189
	12.4502	-23.4352	-3.9352	12.8252	72.2604
$\bar{\mu}_{.j}$	66.1146	53.3333	59.8333	64.7396	61.1585

As expected, we see from Table 3 that the interaction estimates, the row and column population marginal mean estimates, and the estimate of the overall mean all depend on the design weights. However, the F-test for testing the overall significance of the AB interactions as given in (11)

$$F = MS_H / MS_e = \frac{9560.00/6}{2269.92/20} = \frac{1593.33}{113.50} = 14.04, \tag{16}$$

is independent of the system of design weights. Setting $\alpha = 0.05$, the critical value for the F-test is 2.60, with $v_h = 6$ and $v_e = 20$ degrees of freedom. Since the null hypothesis is rejected, we see that the individual estimates of interaction and hence their interpretation are affected by the weights.

Investigating the interaction estimates in Table 3, except for the situation $W_{ij} = n_{ij} > 0$, each $\hat{\gamma}_{ij}$ is estimating a population parameter that has the following general structure:

$$\gamma_{ij} = \mu_{ij} - \text{marginal row mean} - \text{marginal column mean} + \mu \tag{17}$$

where μ_{ij} is a population cell mean, μ is an overall mean, and the marginal means are defined in (10). This follows from (7) and (8) with Σ-restrictions (3) and UV-restrictions (4), respectively. With W-restrictions (5) and $W_{ij} \propto W_{i+}W_{+j}$, (9) reduces to (15) which has the form of (17). These observations lead to the following general theorem.

Theorem 1. Using Σ-restrictions, UV-restrictions or W-restrictions with $W_{ij} \propto W_{i+}W_{+j}$, the interaction parameters have the residual form given in (17) for appropriately defined row, column, and overall means.

From (9), we see that Theorem 1 is not true for an arbitrary set of nonnegative weights $\{W_{ij}\}$. Even so, W-restrictions with weights $W_{ij} = n_{ij}$ yield a nice solution to the ANOVA problem, Fujikoshi (1993).

Bradu and Gabriel (1974) called contrasts in $\boldsymbol{\Gamma}$ that have the simple residual form given by (17) a product-type contrast. Using the matrix $\boldsymbol{\Gamma}$ or the matrix of means $\mathbf{U}$, a product-type contrast is defined as $\psi = \mathbf{a}'\boldsymbol{\Gamma}\mathbf{b} = \mathbf{a}'\mathbf{U}\mathbf{b}$ where the elements of $\mathbf{a}$ and $\mathbf{b}$ are contrasts, the elements of $\mathbf{a}$ and $\mathbf{b}$ sum to zero ($\sum_i a_i = \sum_j b_j = 0$). Even though the interactions do not have the residual form given in (17) for arbitrary W-restrictions, we have the following general result.

Theorem 2. For γ_{ij} defined as the residual $\gamma_{ij} = \mu_{ij} - \mu - \alpha_i - \beta_j$, $\psi = \mathbf{a}'\mathbf{U}\mathbf{b} = \mathbf{a}'\boldsymbol{\Gamma}\mathbf{b}$ for all contrast vectors $\mathbf{a}$ and $\mathbf{b}$ and its value does not depend on Σ-, UV- or W-restrictions.

More generally, Bradu and Gabriel (1974), Milliken and Johnson (1992, p. 116) and Boik (1993) define interaction contrasts as

$$\psi = \sum_{i=1}^{a}\sum_{j=1}^{b} c_{ij}\mu_{ij}, \text{ or equivalently } \psi = \text{trace } (\mathbf{C}'_{AB}\mathbf{U}) \tag{18}$$

where $\mathbf{C}_{AB}$ is an $a \times b$ matrix with elements $\{c_{ij}\}$; the elements in each row and each column of $\mathbf{C}_{AB}$ sum to zero (are contrast vectors) and the function Trace $(\cdot)$ is defined as the sum of the diagonal elements of a square matrix. Thus, a product-type contrast is a special case of an interaction contrast in which the coefficient matrix $\mathbf{C}_{AB} = \mathbf{ab}'$, an outer product matrix of rank one. Again, we have the following general result.

Theorem 3. For γ_{ij} defined as the residual $\gamma_{ij} = \mu_{ij} - \mu - \alpha_i - \beta_j$, $\psi =$ Trace $(\mathbf{C}'_{AB}\mathbf{U}) =$ Trace $(\mathbf{C}'_{AB}\boldsymbol{\Gamma})$ for all contrast matrices $\mathbf{C}_{AB}$ and its value does not depend on Σ-, UV- or W-restrictions.

Thus, product-type contrasts in μ_{ij} or γ_{ij} are a subset of all interaction contrasts.

When one rejects the overall test of interaction using an overall F-statistic, we know from Scheffé (1953; 1959, p. 109) that either an individual interaction term γ_{ij} is significantly different from zero or some parametric function:

$$\theta = \sum_i \sum_j c_{ij}\gamma_{ij} = \text{Trace } (\mathbf{C}'\boldsymbol{\Gamma}) \tag{19}$$

is nonzero for some matrix of coefficients $\mathbf{C}_{a\times b} = \{c_{ij}\}$. Comparing (19) and (18), we see that the matrix $\mathbf{C}_{a\times b}$ is an arbitrary matrix, there are no restrictions on the sum of the elements c_{ij}, while $\mathbf{C}_{AB}$ is a contrast matrix in which the elements in each row and each column sum to zero. There is no contradiction here since by (18) and Theorem 3, all linear combinations θ are contained in ψ. This result is implicit in Boik (1993). To see this, there exists weight matrices $\mathbf{W}_a$ and $\mathbf{W}_b$ such that $\mathbf{W}_a\mathbf{U}\mathbf{W}_b = \boldsymbol{\Gamma}$. Hence, $\mathbf{C}'\mathbf{W}_a\mathbf{U}\mathbf{W}_b = \mathbf{C}'\boldsymbol{\Gamma}$ for an arbitrary matrix $\mathbf{C}$. However, the Trace $(\mathbf{C}'\boldsymbol{\Gamma}) =$ Trace $(\mathbf{C}'\mathbf{W}_a\mathbf{U}\mathbf{W}_b) =$ Trace $(\mathbf{W}_b\mathbf{C}'\mathbf{W}_a\mathbf{U}) =$ Trace $(\mathbf{C}'_{AB}\mathbf{U}) =$ Trace $(\mathbf{C}'_{AB}\boldsymbol{\Gamma})$ since $\mathbf{W}_b\mathbf{C}'\mathbf{W}_a$ is a contrast matrix. This

establishes the result. For the equal weight case, $\mathbf{W}_a$ and $\mathbf{W}_b$ have a simple form. In general, their construction is more complicated, Fujikoshi (1993).

When performing a post-hoc analysis following a test of interaction, one first evaluates which, if any, of the individual interaction parameters, a special case of product contrasts, are significantly different from zero as recommended by Rosnow and Rosenthal (1989a, 1989b) and Boik (1979, 1993). However, one may have to look beyond the individual γ_{ij} to establish significance. From Table 3, the estimates of the interaction terms are different, depending on the design weights chosen. Hence, their interpretation depends on the weights used in the study.

To determine whether an individual interaction γ_{ij} or linear combination $\psi = \text{Trace}\,(\mathbf{C}'_{AB}\boldsymbol{\Gamma}) = \text{Trace}\,(\mathbf{C}'_{AB}\mathbf{U})$ of the interactions is significantly different from zero following a significant overall F-test, one may employ the S-method, Scheffé (1953). The simultaneous $(1-\alpha)$ confidence intervals are given by

$$\hat{\psi} - S\hat{\sigma}_{\hat{\psi}} \quad \le \psi \le \quad \hat{\psi} + S\hat{\sigma}_{\hat{\psi}} \tag{20}$$

where $\hat{\psi}$ is an unbiased estimator of ψ, $\hat{\sigma}_{\hat{\psi}}$ is its estimated standard error and S^2 is the critical constant

$$S^2 = v_h F^{1-\alpha}_{v_h, v_e} \tag{21}$$

where $v_h = (a-1)(b-1)$ and $v_e = N - ab$, for our example. To evaluate (20) for our example, one merely estimates $\hat{\psi}$ and $\hat{\sigma}_{\hat{\psi}}$ with $S = \sqrt{(2)(3)(2.60)} = 3.95$. Boik (1993) provides a general matrix formula for obtaining the variance of an interaction contrast estimator ψ.

For our example, the only solutions that are applicable are the "equal" weight and "balanced" weight cases. For the equal weight case, using (7), the estimated variance of $\hat{\gamma}_{ij}$ is

$$\hat{\sigma}^2_{\hat{\gamma}_{ij}} = \hat{\sigma}^2\left[\left(\frac{ab-a-b+1}{ab}\right)^2 + \left(\frac{1-a}{ab}\right)^2 \sum_{\substack{j'=1 \\ j'\neq j}}^{b} \frac{1}{n_{ij'}} + \left(\frac{1-b}{ab}\right)^2 \sum_{\substack{i'=1 \\ i'\neq i}}^{a} \frac{1}{n_{i'j}} + \frac{1}{(ab)^2} \sum_{\substack{i'=1 \\ i'\neq i}}^{a} \sum_{\substack{j'=1 \\ j'\neq j}}^{b} \frac{1}{n_{i'j'}}\right] \tag{22}$$

To obtain the estimated variance for the balanced weight case, equation (15) is used. That is,

$$\hat{\gamma}_{ij} = y_{ij\cdot} - \tilde{y}_{i\cdot\cdot} - \tilde{y}_{\cdot j\cdot} + \tilde{y}_{\cdot\cdot\cdot} = y_{ij\cdot} - \sum_j W_{ij} y_{ij\cdot} / W_{i+} - \sum_i W_{ij} y_{ij\cdot} / W_{+j} + \sum_i \sum_j W_{ij} y_{ij\cdot} / W_{++} . \tag{23}$$

The estimated standard errors for each $\hat{\gamma}_{ij}$ are given in Table 4. The values of W_{ij} are obtained from Table 1, the sample sizes n_{ij} are shown in Table 2, and from the denominator of the F-statistic, $MS_e = \hat{\sigma}^2 = 113.50$.

Table 4 Estimated Standard Errors, $\hat{\sigma}_{\hat{\gamma}_{ij}}$

Equal Weight Case ($W_{ij} = 1$)

4.556	4.434	4.996	4.319
4.950	4.705	4.733	4.896
4.620	4.950	5.055	4.563

Balanced Weight Case ($W_{ij} = W_{i+} W_{+j}$)

4.956	4.426	5.469	3.671
5.461	4.566	5.211	4.160
5.045	4.832	5.550	3.902

Comparing the absolute value of $\hat{\gamma}_{ij}$, $|\hat{\gamma}_{ij}|$, with the critical constant $S\hat{\sigma}_{\hat{\gamma}_{ij}}$ for each interaction $\hat{\gamma}_{ij}$ for both the "equal" weight and the "balanced" weight cases, γ_{12} and γ_{14} [illegible]gnificantly different from zero for both weighting schemes. However, γ_{32} is significant for the equal weight scheme, but not the balanced weight case. The interval is:

$$\begin{array}{lrcl} \text{Equal} & -40.449 & \le \gamma_{32} \le & -1.343 \\ \text{Balanced} & -30.551 & \le \gamma_{32} \le & 5.521. \end{array} \tag{24}$$

Because the definition of the interaction parameters γ_{ij} depend on the weighting scheme, it is not surprising to find that the confidence sets for the γ_{ij} differ in size. Thus, the significance or nonsignificance of a γ_{ij} may be reported differently by two researchers for the same set of data. The only difference is in the design weights selected for the analysis. This fact is often overlooked by applied researchers when discussing the analysis of a CRF-ab design.

In our analysis of the interactions, we chose to investigate the individual γ_{ij} to locate significance following the rejection of the overall *F*-test. Often these individual γ_{ij} are not significant and one must locate the contrast in the γ_{ij}'s or equivalently the μ_{ij}'s that led to the rejection of the overall test. To find the most significant contrast in γ_{ij} (μ_{ij}) that led to rejection of the overall test, the contrast matrix $\mathbf{C}'_{AB}$ must be selected proportional to Γ' for a design with equal cell frequencies $n_{ij} = n$, Hochberg and Tamhane (1987, p. 296). This is the case since $MS_H = SS_H / \nu_h$ and SS_H in (12) for equal $n = n_{ij}$ has the form $SS_H = n \text{ Trace}(\hat{\Gamma}\hat{\mathbf{T}})$. Extending this result to the unbalanced design, we let $W_{ij} = n_{ij}$. Then the most significant (maximum) contrast is

$$\psi = \sum_{i,j} (n_{ij}\hat{\gamma}_{ij})\gamma_{ij} = \text{Trace}[(\mathbf{N} * \hat{\Gamma})'\Gamma] = \text{Trace}(\mathbf{C}'_{AB}\Gamma) = \text{Trace}(\mathbf{C}'_{AB}\mathbf{U}) \tag{25}$$

where the matrix $\mathbf{N} = \{n_{ij}\}$, the matrix of cell frequencies, and $\mathbf{N} * \Gamma = \{n_{ij}\gamma_{ij}\}$, is a Hadamard product. For $W_{ij} = n_{ij}$, one substitutes $\hat{\gamma}_{ij}$ into (25) to obtain the coefficient matrix $\mathbf{C}'_{AB}$ for ψ. The value of $\hat{\psi} = 9560.00 = SS_H$, as expected. Using (20), the confidence interval for ψ in (25) is

$$\begin{array}{rcl} \hat{\psi} - S\hat{\sigma}_{\hat{\psi}} & \le \psi \le & \hat{\psi} + S\hat{\sigma}_{\hat{\psi}} \\ 9560 - (3.95)(1041.64) & \le \psi \le & 9560 + (3.95)(1041.64) \\ 5445.85 & \le \psi \le & 13674.18 \end{array} \tag{26}$$

which is significant, does not depend on the design weights, but is impossible to interpret.

Another class of contrasts that are often studied following the F-test are called tetrad contrasts. Tetrad contrasts involve four cells in the design, may be generated from the matrix $\mathbf{\Psi} = \mathbf{C}_A'\mathbf{U}\mathbf{C}_B$, where $\mathbf{C}_A'$ and $\mathbf{C}_B$ are simple contrast matrices, and do not depend on the design weights. To illustrate, we let

$$\mathbf{C}_A' = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \text{ and } \mathbf{C}_B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{pmatrix} \tag{27}$$

for our example. Then $\mathbf{C}_{AB} = \mathbf{c}_{A(i)}\mathbf{c}_{B(i)}'$ where $\mathbf{c}_{A(i)}$ is the i^{th} column of $\mathbf{C}_A$ and $\mathbf{c}_{B(j)}$ is the j^{th} column of $\mathbf{C}_B$ so that $\psi_{ij} = \mathbf{c}_{A(i)}'\mathbf{U}\mathbf{c}_{B(i)} = \text{Trace}(\mathbf{C}_{AB}'\mathbf{U}) = \text{Trace}(\mathbf{C}_{AB}'\mathbf{\Gamma})$ is a tetrad contrast. Thus, tetrad contrasts are a subset of all product contrasts.

For the matrices $\mathbf{C}_A$ and $\mathbf{C}_B$ defined in (27), the six tetrad contrasts are shown in Table 5. In Table 5, observe that only one of the tetrad contrasts is significantly different from zero using the S-method, for our example.

Table 5 Select Tetrad Contrasts Using the S-Method

ψ	$\hat{\psi}$	$\hat{\sigma}_{\hat{\psi}}$	(Sig)	Low Limit	Upper Limit
$\gamma_{11} - \gamma_{31} - \gamma_{14} + \gamma_{34}$	27.250	11.911		-19.794	74.294
$\gamma_{12} - \gamma_{32} - \gamma_{14} + \gamma_{34}$	88.917	12.680	*	38.834	138.999
$\gamma_{13} - \gamma_{33} - \gamma_{14} + \gamma_{34}$	26.917	13.405		-26.030	79.863
$\gamma_{21} - \gamma_{31} - \gamma_{24} + \gamma_{34}$	-31.500	13.754		-85.822	22.822
$\gamma_{22} - \gamma_{32} - \gamma_{24} + \gamma_{34}$	9.333	13.754		-44.989	63.655
$\gamma_{23} - \gamma_{33} - \gamma_{24} + \gamma_{34}$	21.333	13.754		-32.989	75.655

Having selected the design weights for a two-way CRF-ab design, the S-method is the most appropriate simultaneous test procedure for evaluating the significance of an arbitrary number of linear combinations of the γ_{ij}. If one a-priori restricts their investigation to only product-type contrasts, the maximal F-test should be calculated to perform the overall test of significance since the procedure is more powerful than the F-test, Boik (1993). The maximal F-test uses the Studentized Maximum Root (SMR) distribution, the critical value for a two-way design is $R_{p,q,\nu_e}^{1-\alpha}$ where $p = \min(a-1, b-1)$, $q = \max(a-1, b-1)$ and $\nu_e = N - ab$. When $p = 1$, $R_{p,q,\nu_e}^{1-\alpha} = qF_{q,\nu_e}^{1-\alpha}$ reduces to the F-distribution.

For our example, $p = 2$ and $q = 3$. For $\alpha = 0.05$, the critical value of the SMR distribution is $R_{2,3,20}^{0.95} = 13.221$ so that $R = 3.64 < S = 3.95$. Hence, using (20) and substituting R for S, we see that the confidence set for product contrasts will always be shorter, and hence more resolute. Boik (1993) recommends the procedure be used when one is only interested in product-type contrasts, which include for example the individual γ_{ij} and tetrad contrasts. Again, the $\hat{\gamma}_{ij}$ depend on the weighting scheme.

Using a full rank cell means model, Boik (1993) developed SAS (SAS Institute, 1990) and SPSS (1990) programs to calculate the maximal F-statistic. Using the program for our example, the maximal F-statistic is:

$$\text{Maximal } F = \frac{(\mathbf{a}'\hat{\Gamma}\mathbf{b})^2}{\hat{\sigma}^2_{\hat{\psi}}} = \frac{(\mathbf{a}'\mathbf{U}\mathbf{b})^2}{\hat{\sigma}^2_{\hat{\psi}}} = \frac{(51.805)^2}{6.257} = 68.55 \tag{28}$$

where the maximal contrast vectors are:

$$\mathbf{a} = \begin{pmatrix} 0.8156 \\ -0.4406 \\ -0.3750 \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} 0.1273 \\ 0.7333 \\ -0.2358 \\ -0.6248 \end{pmatrix} \tag{29}$$

Comparing the maximal F-ratio to the SMR critical value $R^{0.95}_{2,3,20} = 13.221$, the interaction hypothesis is rejected. The maximal product contrast, similar to the maximum generalized contrast given in (26), is also difficult to interpret. However, using it as a guide, a product contrast that may be more meaningful is $\psi = \mathbf{a}'\mathbf{U}\mathbf{b}$ where $\mathbf{a}' = (1, -1/2, -1/2)$ and $\mathbf{b}' = (0, 1, 0, -1)$ which compares A_1 with the average of A_2 and A_3 for the levels of factor B at B_2 and B_4. For this contrast, $\hat{\psi} = 84.25$, $\hat{\sigma}_{\hat{\psi}} = 10.653$ and $\hat{\psi} / \hat{\sigma}_{\hat{\psi}} = 7.909 > R = 3.64$ so that the comparison is significant. While one may continue to "data-snoop" among the product contrasts, including the individual γ_{ij}, to locate significant interactions that may be meaningful, the simplest and most easily interpreted contrasts are again the tetrad contrasts. Using the maximal F-test, the intervals shown in Table 5 would be more resolute. However, can we do better? The answer is yes. If a researcher is only interested in the $\binom{a}{2}\binom{b}{2}$ tetrad contrasts, the finite intersection test (FIT) procedure developed by Krishnaiah (1964, 1965) yields the shortest confidence sets. Furthermore, all tetrad contrasts are easily interpreted and they are weight invariant. The FIT procedure uses the multivariate F-distribution and requires the use of a computer program to approximate its critical values Cox et al. (1994) and Timm (1995).

To evaluate the significance or nonsignificance of the $\binom{3}{2}\binom{4}{2} = 18$ tetrad contrasts for our example, the statistics $T_i^2 = \hat{\psi}_i^2 / \hat{\sigma}^2_{\hat{\psi}_i}$ are calculated and each is compared to the critical value of the multivariate F-distribution with 1 and 20 degrees of freedom. For $\alpha = 0.05$, Šidák's upper product bound (see e.g. Krishnaiah, 1979 or Cox et al., 1980 for a discussion of calculating Šidák's bound) for the multivariate F-distribution is 11.266. Thus, to establish $1-\alpha$ simultaneous confidence intervals the critical constant

$$FIT = \sqrt{11.266} = 3.36 < R = 3.63 < S = 3.95 \tag{30}$$

is always less than the corresponding critical constants for the SMR distribution or the F-distribution. The FIT is uniformly shorter if one is only interested in tetrad contrasts.

To use the FIT program, the cell means for our example are arranged into a vector:

$$\mu' = (\mu_{11}, \mu_{12}, \mu_{13}, \mu_{14}, \mu_{21}, \mu_{22}, \mu_{23}, \mu_{24}, \mu_{31}, \mu_{32}, \mu_{33}, \mu_{34}) \tag{31}$$

and a contrast matrix for the 18 tetrad contrasts is input to the FIT program:

$$\underset{(18\times12)}{\mathbf{C}'} = \begin{pmatrix}
1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & -1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & -1 & 1 \\
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\
1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\
1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\
0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\
0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\
0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1
\end{pmatrix}$$

Letting $\Psi = \mathbf{C}'\mu$, the estimate of ψ_i is $\hat{\psi}_i = \mathbf{c}_i'\hat{\mu}$ where $\hat{\mu}$ is the vector of cell means, ψ_i and $\mathbf{c}_i'$ are the i^{th} row of Ψ and $\mathbf{C}'$, respectively. The output for the FIT program is provided in Table 6.

Table 6 FIT Output for (3×4) Design

TWO SIDED FINITE INTERSECTION TEST
(* INDICATES TO REJECT THE SUBHYPOTHESIS OF NO DIFFERENCE)

VARIABLE 1:		SCORE
DEGREES OF FREEDOM	1, 20	
S-SQUARE/NDF:	113.4960	
		SIDAK'S UPPER BOUND ON MULTIVARIATE F
CRITCAL VALUE:		11.266
LEVEL OF SIGNIFICANCE FOR VARIABLE 1:		0.050

LINEAR COMBINATION	STATISTIC	ACC/REJ	LINEAR COMBINATIONS OF THE ORIGINAL MEANS ESTIMATE	CONFIDENCE	INTERVAL
1	2.549		20.8333	[-64.6277,	22.9610]
2	14.943	*	53.1667	[7.0033,	99.3300]
3	19.207	*	58.7500	[13.7556,	103.7444]
4	32.166	*	74.0000	[30.2056,	117.7944]
5	39.391	*	79.5833	[37.0229,	112.1438]
6	0.173		5.5833	[-39.4111,	50.5778]
7	8.815		-40.8333	[-86.9967,	5.3300]
8	14.757	*	-52.8333	[-98.9966,	-6.6700]
9	5.246		-31.5000	[-77.6633,	-14.6633]
10	0.761		-12.0000	[-58.1633,	-34.1633]
11	0.461		9.3333	[-36.8300,	55.4966]
12	2.406		21.3333	[-24.8300,	67.4966]
13	22.337	*	-61.6667	[-105.4610,	-17.8723]
14	0.001		0.3333	[-45.8300,	46.4967]
15	5.234		27.2500	[-12.7286,	67.2286]
16	18.474	*	62.0000	[13.5835,	110.4165]
17	49.172	*	88.9167	[46.3562,	131.4771]
18	4.032		26.9167	[-18.0778,	71.9111]

clearly shows the significant tetrad contrasts for the Overall and Spiegel data. Comparing entries 15, 17, 18, 9, 11, and 12 in this table with those in Table 5, observe that the tetrad confidence intervals for the FIT procedure are shorter in all cases. Given the relationship among the critical constants in (30), the FIT procedure also produces confidences intervals that would be shorter than those realized using the maximal F-criterion. Hence, if one is only interested in tetrad contrasts, the FIT procedure should be utilized in the analysis of interactions for an $a \times b$ design. Of course the FIT procedure is not limited to tetrad contrasts. It may be used with any finite number of contrasts.

To perform a step-down FIT procedure for this example, one would remove tetrad 17, $\psi = \mu_{12} - \mu_{14} - \mu_{32} + \mu_{34}$ and recalculate the overall multivariate F-critical value, continuing to remove the tetrad corresponding to the largest statistic at each step and stopping the process when nonsignificance is realized. For the example, the step-down procedure found one more tetrad to be significant, tetrad number 15. The sequence of multivariate F-critical values was:

{11.266, 10.946, 10.771, 10.586, 10.388, 10.176, 9.946, 9.697, 9.424}

for the step-down FIT procedure. The step-down process will always yield at least the number of significant contrasts found using the single-step procedure, and maybe more. However, it is difficult to establish $1-\alpha$ simultaneous confidence bounds for the population parameters at each step, Timm (1995).

Conclusion

This paper began with a researcher interested in analyzing whether there is a significant interaction between two factors in a completely randomized factorial design. In implementing the study, an unbalanced nonorthogonal design resulted. To analyze the data, one researcher used equal cell weights and the other used proportional cell or balanced weights. In performing the analysis, we found that the interaction parameters γ_{ij} and their corresponding estimates $\hat{\gamma}_{ij}$ depend on the weighting scheme selected. Hence,

when constructing confidence intervals for γ_{ij} or parametric functions of the γ_{ij} one must discuss the weights used in the analysis.

While this *F*-test and corresponding S-method is used to investigate all γ_{ij} and contrasts in the γ_{ij}, a more powerful test exists if one restricts their investigation of interactions to γ_{ij} and only product contrasts in the γ_{ij}. The most powerful overall test is the maximal *F*-test. When constructing confidence sets for the γ_{ij}, we saw that they still depend on the weighting scheme for the analysis. If a researcher in the study of interactions is only interested in all tetrad contrasts of the γ_{ij}, no overall test is performed. Instead, one uses the finite intersection test (FIT) procedure for the analysis. The overall test of significant interaction is rejected if any tetrad is significantly different from zero. This approach is particularly attractive in the analysis of interactions for unbalanced designs since the procedure does not depend on the weighting scheme selected.

In performing an analysis of interactions in unbalanced factorial designs, researchers have a responsibility for reporting the design weights when analyzing interactions γ_{ij}. If one is only interested in tetrad contrasts, the FIT procedure yields confidence sets that have the smallest probability of covering zero and hence are more likely to yield significant results.

References

Appelbaum, M.I., & Cramer, E.M. (1974). Some problems in the nonorthogonal analysis of variance. Psychological Bulletin, 81, 335-343.

Arnold, A.F. (1981). The theory of linear models and multivariate analysis. New York: John Wiley.

Boik, R.J. (1979). Interactions, partial interactions, and interaction contrasts in the analysis of variance. Psychological Bulletin, 86, 1084-1089.

Boik, R.J. (1993). The analysis of two-factor interactions in fixed effects linear models. Journal of Educational Statistics, 18, 1-40.

Bradu, D., & Gabriel, K.R. (1974). Simultaneous statistical inference on interactions in two-way analysis of variance. Journal of the American Statistical Association, 69, 428-436.

BMDP (1992). BMDP statistical software manual. Vol. 1 & 2, (Release 7.0 Ed.). W.J. Dixon, editor. Los Angeles, CA: University of California Press.

Carlson, J.E., & Timm, N.H. (1974). Analysis of nonorthogonal fixed-effects designs. Psychological Bulletin, 81, 563-570

Cohen, J. (1988). Statistical power analysis for behavioral science (2nd ed.). Hillside, NJ: Erlbaum.

Cox, C.M., Krishnaiah, P.R., Lee, J.C., Reising, J. and Schuurman (1980). A study of finite intersection tests for multiple comparisons of means. Multivariable analysis-V (P.K. Krishnaiah, Ed.), 435-466. New York: Academic Press.

Cox, C.M., Fang, C., Boudreau, R.M., & Timm, N.H. (1994). Computer program for Krishnaiah's finite intersection tests for multiple comparisons of mean vectors, Interim Report No. 94-04, Research Methodology Program, University of Pittsburgh, Pittsburgh, PA.

Cramer, E.M., & Appelbaum, M.I. (1980). Nonorthogonal analysis of variance—Once again. Psychological Bulletin, 87, 51-57.

Davidson, M., & Toporek, J. (1991). BMDP 4V - General univariate and multivariate analysis of variance. BMDP Technical Report No. 67. Los Angeles, CA: BMDP Statistical Software, Inc.

Fujikoshi, Y. (1993). Two-way ANOVA models with unbalanced data, Discrete Mathematics, 116, 315-334.

Graybill, F.A. (1976). Theory and application of the linear model. Boston: Duxburg Press.

Hockberg, Y., & Tamhane, A.C. (1987). Multiple comparison procedures. New York: Wiley.

Keren, G., & Lewis, C. (1977). A comment on coding in nonorthogonal designs. Psychological Bulletin, 84, 346-348.

Kirk, R.E. (1995). Experimental design: Procedure for the behavioral sciences, (3rd Edition). Pacific Grove, CA: Brooks/Cole.

Krishnaiah, P.R. (1964). Multiple comparison tests in multivariate case. ARL 64-124. Wright-Patterson Air Force Base, Ohio.

Krishnaiah, P.R. (1965). On the simultaneous ANOVA tests. Annals of the Institute of Statistical Mathematics, 17, 167-173.

Krishnaiah, P.R. (1979). Some developments on simultaneous test procedures. Developments in statistics, (P.R. Krishnaiah, Ed.), Vol. 2, 157-201. New York: Academic Press.

Lewis, C., & Keren, G. (1977). You can't have your cake and eat it too: Some considerations of the error term. Psychological Bulletin, 84, 1150-1154.

Milliken, G.A., & Johnson, D.E. (1992). Analysis of messy data. New York: Chapman and Hall.

O'Brien, R.G. (1976). Comments on "Some problems in the nonorthogonal analysis of variance." Psychological Bulletin, 83, 72-74.

Overall, J.E., & Spiegel, D.K. (1969). Concerning least squares analysis of experimental data. Psychological Bulletin, 72, 311-322.

Overall, J.E. & Spiegel, D.K. (1973). Comment on "Regression analysis of proportional data." Psychological Bulletin, 80, 28-30.

Overall, J.E., Spiegel, D.K., & Cohen, J. (1975). Equivalence of orthogonal and nonorthogonal analysis of variance. Psychological Bulletin, 82, 182-186.

Rawlings, R.R., Jr. (1972). Note on nonorthogonal analysis of variance. Psychological Bulletin, 77, 373-374.

Rawlings, R.R., Jr. (1973). Comments on the Overall and Spiegel paper. Psychological Bulletin, 79, 168-169.

Rosnow, R.L., & Rosenthal, R. (1989a). Definition and interpretation of interaction effects. Psychological Bulletin, 105, 143-146.

Rosnow, R.L., & Rosenthal, R. (1989b). Statistical procedures and the justification of knowledge in psychological science. American Psychologist, 44, 1276-1284.

SAS Institute (1990). SAS/STAT user's guide (Version 6, 4th ed). Cary, NC: Author

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance. Biometrika, 40, 87-104.

Scheffé, H. (1959). The analysis of variance. New York: Wiley.

Searle, S.R. (1971). Linear models. New York: Wiley.

Searle, S.R. (1993). Unbalanced data and cell means models. Applied analysis of variance in behavioral science (L.K. Edwards, Ed.), 375-420. New York: Marcel Dekker.

SPSS. (1990). SPSS reference guide. Chicago: Author.

Timm, N.H., & Carlson, J.E. (1975). Analysis of variance through full rank models. Multivariate Behavioral Research, Monograph, 1975, No. 75-1.

Timm, N.H. (1995). Simultaneous inference using finite intersection tests: A better mousetrap. Multivariate Behavioral Research, 30 (4), 461-512.